DEFENCE R&D DÉFENSE

# Measurement accuracy and precision assessment of an image-based clothing and equipment sizing system


*P. Meunier*
*S. Yin*



DEFENCE AND CIVIL
INSTITUTE OF ENVIRONMENTAL MEDICINE



Technical Report
DCIEM TR 1999-070
July 1999


National Defence Défense nationale

Canada

19990928 397

July 1999

DCIEM TR1999-070

# MEASUREMENT ACCURACY AND PRECISION ASSESSMENT OF AN IMAGE-BASED CLOTHING AND EQUIPMENT SIZING SYSTEM

P. Meunier
S. Yin

Defence and Civil Institute of Environmental Medicine
1133 Sheppard Avenue West, P.O. Box 2000
Toronto, Ontario
Canada M3M 3B9



DEPARTMENT OF NATIONAL DEFENCE – CANADA

## EXECUTIVE SUMMARY

In spite of highly standardised protocols designed to maximise the degree of repeatability and accuracy, traditional anthropometric data are not as reliable as they appear. Many factors come into play during the physical measurement of human subjects, resulting in numerous possible sources of error. Researchers have found the magnitude of these errors to be such that, even if measured by highly trained observers, comparison of two populations may be meaningless.

Computerised image-based systems can overcome some of the problems of traditional anthropometry, such as error due to instrument alignment, the pressure exerted on soft tissue by the various measurement instruments, or even transcription errors. However, all sources of error have not been eliminated. In image-based systems, the sources of error take the form of perspective distortion, camera resolution, and inadequacy of the mathematical models used to estimate circumference measurements.

The accuracy of measurements made by an image-based clothing and equipment sizing system was estimated using a database of 349 subjects (male and female) who were also measured traditionally. The precision, or repeatability, of this system was estimated through repeated measurements of both a plastic mannequin and a human. Although the image-based system did not exhibit systematic bias in the results, the standard deviations were somewhat smaller for some dimensions than those obtained by manual measurement. The repeatability results were comparable to those obtained by highly trained anthropometrists, as reported in recent large-scale surveys. The reliability of the measurements needed for clothing, i.e. the proportion of error of measurement to biological variability, was greater than 99% in all cases.

The degree of accuracy and precision of the measurements required for the selection of clothing and equipment size was put into perspective with the realities of short-term fluctuations in body size, clothing design, and manufacturing tolerances. When a balanced approach is used, neck circumference is found to be, by far, the anthropometric dimension requiring the greatest amount of accuracy. Because of the ease with which it can be identified and measured by image processing, it is also the system's most accurately measured circumference.

When properly designed and calibrated, image-based systems can provide unbiased anthropometric measurements that are quite comparable to traditional measurement methods (performed by skilled measurers), both in terms of accuracy and repeatability. The quality of the results depends, in large part, on the dependability of the automatic landmarking algorithms and the correct modelling, but once this is achieved, this type of system can provide a reliable basis for the measurement of a population, regardless of where, when or by whom, it is operated.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

## INTRODUCTION

In spite of highly standardised protocols designed to maximise the degree of repeatability and accuracy, anthropometric data are not always as reliable as they appear. Many factors come into play during the measurement of human subjects, resulting in numerous possible sources of error. Some of the important sources include posture, identification of landmarks, instrument position and orientation, and pressure exerted by the measuring instrument, to name a few (Davenport *et al.*, 1935). In fact, it has been said that true values are seldom measured in anthropometry (Jamison & Zegura, 1974).

Researchers have found the magnitude of these errors to be such that, even if measured by highly trained observers, comparison of two populations may be meaningless (Bennett & Osborne, 1986). In a comparative study by Kemper & Pieters (1974), fifty boys (12 and 13 years of age) were measured independently by experienced observers in two institutes. Both teams of observers were trained to the same measurement techniques and used the same measuring instruments. In spite of this, systematic differences were found in 9 of the 12 measurements taken. Pearson correlation coefficients between 0.872 (biacromial diameter) and 0.996 (stature) were found for the measurements taken by the two groups. Although the lowest correlation (biacromial diameter) did not present systematic errors, it suffered from repeatability problems (precision error).

In another study of anthropometric inter-observer error, Jamison & Zegura (1974) compared the measurements made by two anthropometrists on the same group of 42 individuals (20 males and 22 females). The same instructor had trained both anthropometrists at the same time. The results, which were analysed univariately and multivariately, showed a significant degree of systematic bias between the observations. Only 5 out of 16 measurements had correlations higher than 0.90, which can be interpreted as meaning that only 81% ($r^2 = 0.90^2$) of the variability is accounted for. The results of these and many more studies show how difficult it is to measure humans, even under controlled conditions and after extensive training of the observers.

Computerised image-based systems, such as the Intelligent Clothing and Equipment Sizing System (ICESS), can overcome some of the problems of traditional anthropometry. For instance:

- specialised training of observers is not required, since the computer contains all of the expertise required;
- image processing and shape recognition algorithms can repeatably identify key body shape features;
- measurements are not biased by pressure exerted on soft tissue;
- reading and transcription errors are eliminated.

All errors are not eliminated, however, as is the case for any measurement system. In the case of ICESS, the sources of error take the form of perspective distortion, camera resolution, and inadequate models for circumference measurements. The objective of this paper is to evaluate the accuracy of the measurements made by ICESS, and put them in perspective with traditional anthropometry and the clothing application it was designed for.

# BACKGROUND

## Error

The error of a measurement is defined as the difference between the measured value and the true value of the item being measured. Errors can be catalogued as either random (precision error) or systematic (bias error). *Precision* is defined as the difference in values obtained when measuring the same object repeatedly. It has an average value of zero. *Accuracy* is the difference between the measured and true values. Bias error, which occurs in the same way on each measurement, affects the accuracy of a measurement while random error affects precision. The result of both types of error is called uncertainty, and is defined in the following way (Beckwith *et al.*, 1993):

$$U_x = (B_x^2 + P_x^2)^{1/2}$$

Where B is the bias and P is the precision, both of which should have the same confidence level, i.e. 95%.

This concept of error is useful, but it relies on knowledge of the true value of what is being measured. Since any measurement contains error, the pure error can not be calculated. However, it can be estimated. Precision error can be estimated by taking a large number of readings on an individual and by using a statistical model to determine the expected spread of values at a given probability level. Bias error, on the other hand, requires comparison of measurements with a more accurate method/instrument. This is difficult to do in anthropometry, given that the best available method is one that contains non-negligible error itself.

## ICESS

### System description

ICESS is a PC-based system comprised of two Kodak DC120 colour digital cameras (1280 x 960 pixels) and a blue backdrop embedded with calibration markers (Figure 1). The system takes simultaneous (within a fraction of a second) front and side pictures of individuals standing with their arms alongside slightly abducted. By taking both images simultaneously, the exact posture in space is captured, and it is possible to recover the object's three-dimensional size.

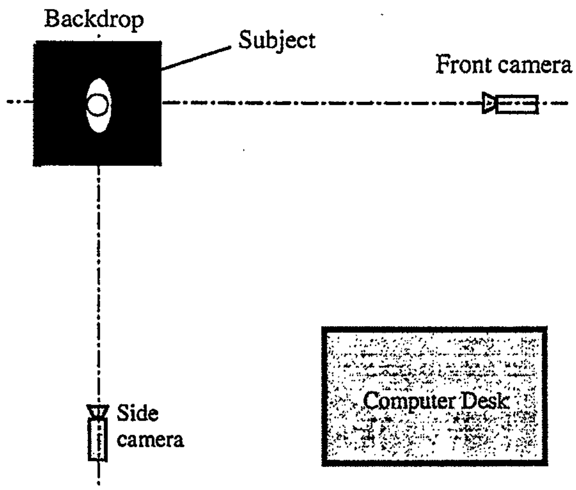


Figure 1 Plan view of ICESS setup.

The image analysis process, illustrated in Figure 2, requires a) pre-processing of the images, b) calibration of the cameras, c) segmentation of the body from the background, d) landmark detection, and e) calculation of the anthropometric variables.

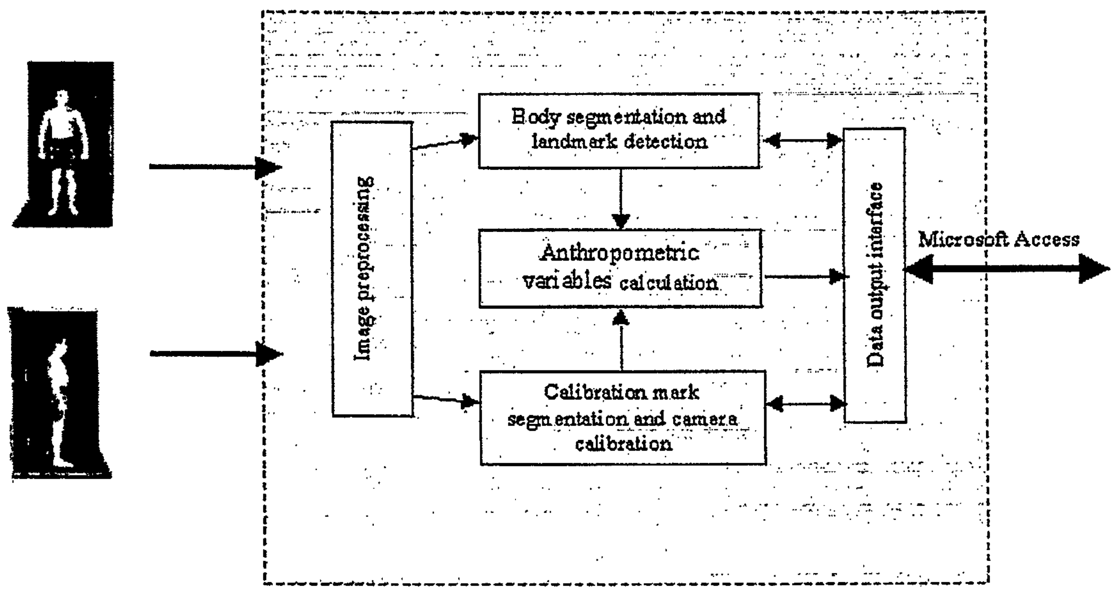


Figure 2 Image analysis process.

Potential sources of error can be found at each of these steps. The following is a short discussion of these sources and what was done in ICESS to mitigate their effect:

a) Image pre-processing is required in order to remove image noise. This noise may come from the image sensor pixel, the analog-to-digital converter, uncontrolled lighting, or even dust between the camera and the viewing object. These random sources of noise can cause instability of some of the image processing algorithms. There are a number of image data restoration algorithms. However, there is always a compromise between the amount noise reduction and the loss of useful information. A medial filter (Schalkoff, 1989) was selected for ICESS because of its better structure preservation qualities.

b) Camera calibration is required to calculate camera parameters such as the focal length, the position and orientation (relative to the object co-ordinates), the optical axis centre, and the lens distortion. Calibration errors were minimised by relying on the accurate positioning of the calibration markers on a solid substrate. The markers are inextricably linked to the image so that inadvertent movement of the cameras between pictures is of no consequence, and to allow re-processing to be performed.

c) Segmentation of the individual from the background is one of the factors affecting measurement accuracy. For this reason, special attention was paid to how segmentation was performed. An adaptive, multi-pass segmentation process was designed for ICESS in which the general body features were identified in the initial pass, followed by a more refined segmentation based on this knowledge.

d) ICESS does not require the prior landmarking of subjects. It is able to detect the location of landmarks automatically using shape information. Identification of the landmarks is highly dependent on shape, and as such its accuracy will vary according to where the measurements need to be taken. The use of colour allowed landmarks such as crotch height and chest breadth to be identified. These are often obscured in high contrast or shadowed images. Much effort was spent on landmark identification algorithms that were robust enough to work for all body shapes and sizes.

e) Lengths, breadths and depths are measured directly by the system, while circumferences are obtained indirectly through modelling. Direct measurements are not sources of error in and of themselves, but rather a reflection of the errors injected in the previous four steps. Indirect measurements, on the other hand, require the combination of direct measurement using a mathematical model. They are therefore subject to direct measurement error, model errors as well as errors from the four previous steps.

Theoretical assessment of error

An estimate of measurement error can be made from a theoretical perspective, using the camera resolution as the starting point. The cameras used in ICESS have 1280 by 960 pixels covering an area that is approximately 2.5 m by 1.8 m at the subject. This corresponds to a resolution of 2.0 mm/pixel. Assuming segmentation error of plus or minus one pixel, direct measurements requiring two points (i.e. for breadths, depths, and heights) will likely fluctuate within ± 2 mm (2 pixels x 2 mm/pixel) of the true value. The maximum error, which is obtained when both points err in making the dimension too small or too large, would put the result within ± 4 mm of the true value (2 pixels x 2 mm/pixel).

Circumferences can not be measured directly using only front and side pictures. They must be estimated from direct measurements of breadth and depth using mathematical models. The choice of model depends on the cross-sectional shape being measured. Assuming a perfect model, as is the case for the measurement of a cylindrical object, the maximum error will occur when both breadth and depth measurements err on the same side. The circumference measurement should be within ± 6 mm ($\pi \times (d_1 - d_2) = \pi \times 2$ mm) of the true value for a one pixel error on the circumference, or ± 13 mm for a two pixel error.

Error coming from the model is very difficult to estimate from a theoretical standpoint since it is specific to the shape being measured. For example, an elliptical model may be used to estimate hip circumference using hip breadth and depth as input. Since hips are not usually perfectly elliptical, a certain degree of error can be expected from such a model. This error is in addition to segmentation and resolution error made on the two direct measurements required as input to the model. Empirical data are required to determine the magnitude of this error.

## METHODOLOGY

### Accuracy assessment

The accuracy of the image-based system was assessed by comparing image-based measurements with manual measurements taken by anthropometrists during the 1997 survey of the Canadian Land Forces (Chamberland *et al.*, 1998). Six dimensions were selected because of their relevance to clothing sizing, which is the main purpose of the system. These were: stature, neck circumference, chest circumference, waist circumference, hip circumference, and sleeve length (spine-wrist).

The test sample consisted of a subset of 349 subjects (95 females and 254 males) from the survey that had been measured both with traditional methods and with the image-based system. The image capture was performed within 90 minutes of the traditional measurements to avoid the effects of daily body variations. T-tests were performed to compare the means of all dimensions. Waist circumference was excluded from this comparison due to the difference in measurement definition between the two methods.

### Precision assessment

The precision of the image-based system was determined by performing repeated measurements on a full size plastic mannequin as well as on a human subject. All image capture and analysis sequences were performed in succession (every minute or so) such that camera calibration and lighting conditions were relatively constant. The mannequin was used in order to exclude variations due to breathing movement and postural differences from picture to picture. The subject was instructed to stand with the arms slightly abducted along the side the body during picture taking, and to move away from the platform between measurements. Thus, the precision estimates obtained this way contain variability coming from postural differences, breathing movement, and repositioning from one set of images to the other.

## RESULTS

### Accuracy

ICESS currently performs over 30 anthropometric measurements. Six of these dimensions are of particular interest in clothing sizing, which is the main purpose of the system. These are: stature, neck circumference, chest circumference, waist circumference, hip circumference, and sleeve length (spine-wrist). A detailed analysis of the performance of ICESS compared to the manual measurements taken during the LF97 survey was performed on those measurements, with the exception of waist circumference. Waist circumference is unique in that the landmarks used by ICESS for clothing purposes are different than those used in the LF97 survey. ICESS measures waist circumference where trousers/slacks are worn, whereas the LF97 survey used anatomically defined landmarks such as omphalion. Because of this difference, the two measurements are not equivalent and can not be compared in the same manner as the other measurements.

The results, shown in figures 3 to 14, illustrate the similarity of manual and ICESS measurements. Comparison of the means obtained by those two methods, using t-tests for dependent samples, showed no significant differences. Odd numbered figures show box and whisker plots comparing the means (central dot), standard deviations (top and bottom edges of the box), and the range of 95% of the observations (whiskers). Even numbered figures show scatterplots of the raw results of manual and ICESS measurements, illustrating how well correlated they are. Pearson correlation values, "r", are listed in the legend for each gender.

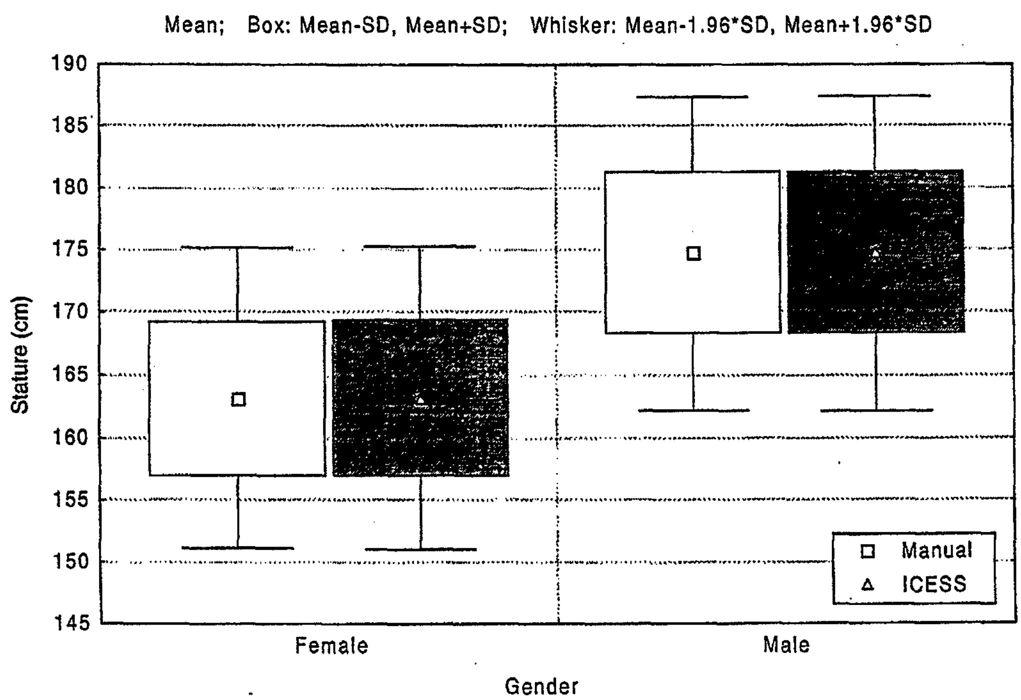


Figure 3 Comparison of manual and ICESS stature measurements.

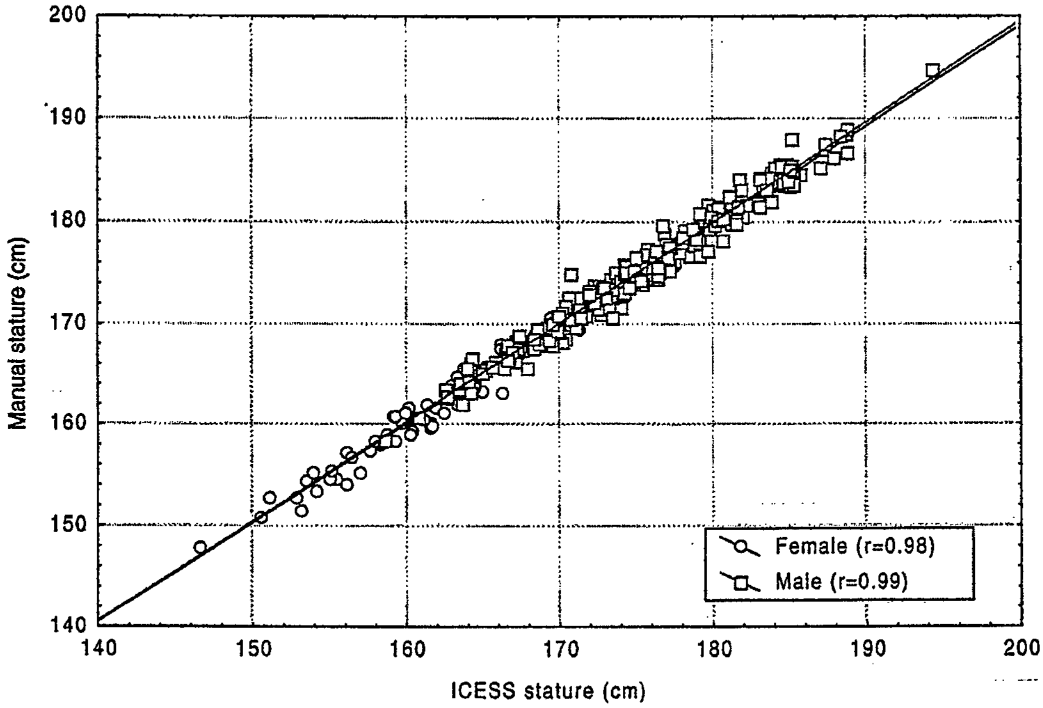


Figure 4 Scatter plot of manual and ICESS stature measurements.

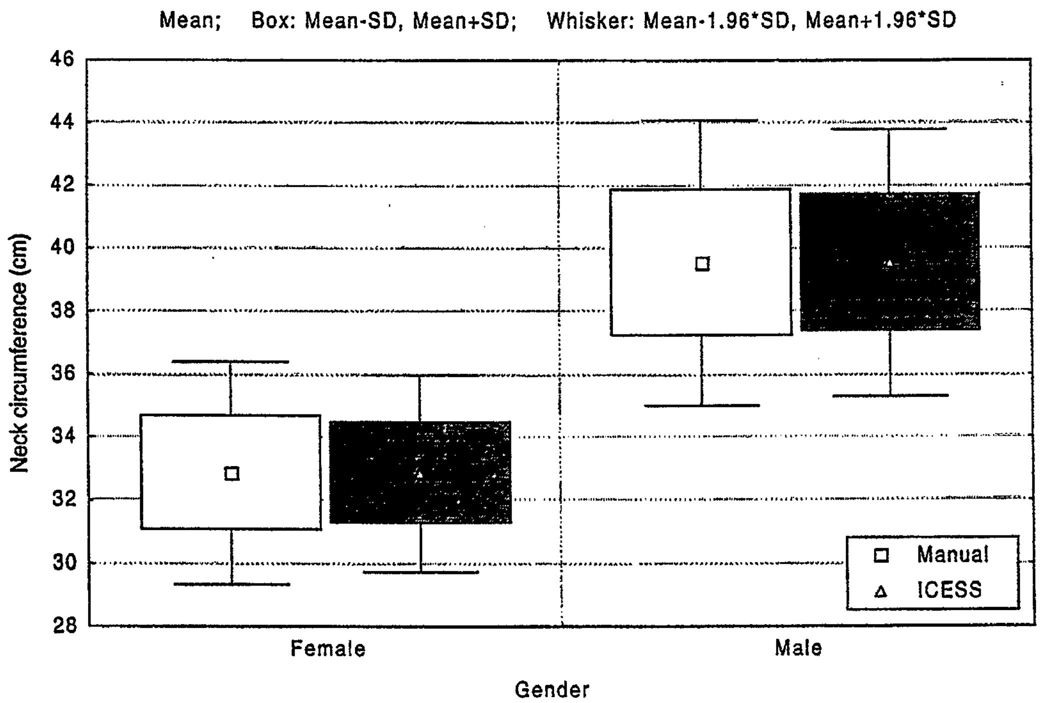


Figure 5 Comparison of manual and ICESS neck circumference measurements.

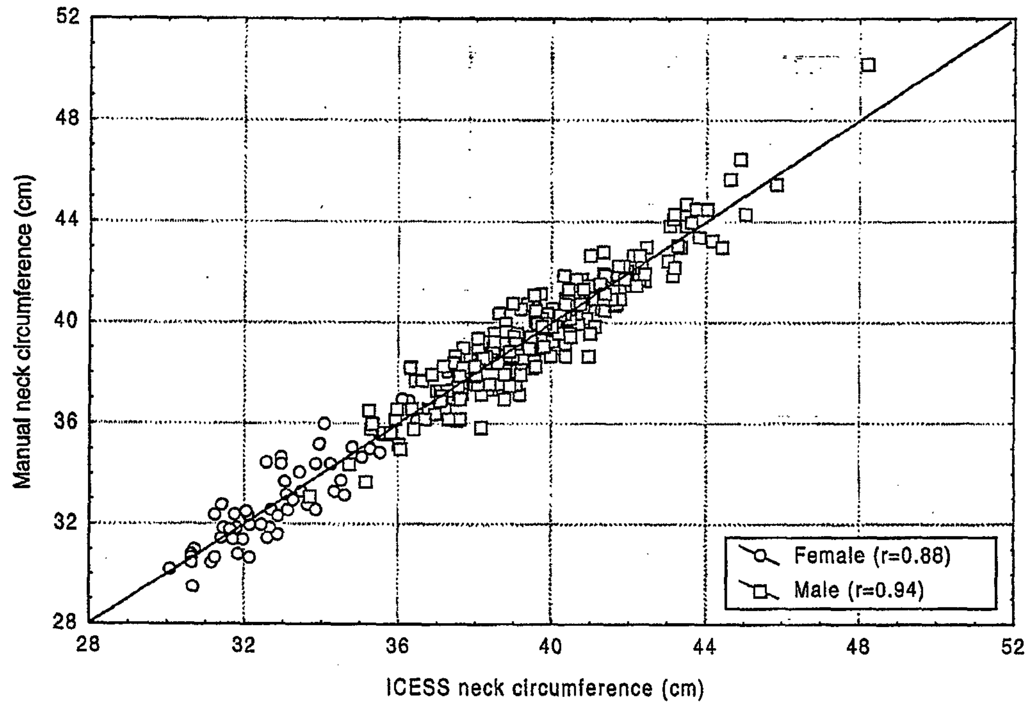


Figure 6 Scatter plot of manual and ICESS neck circumference measurements.

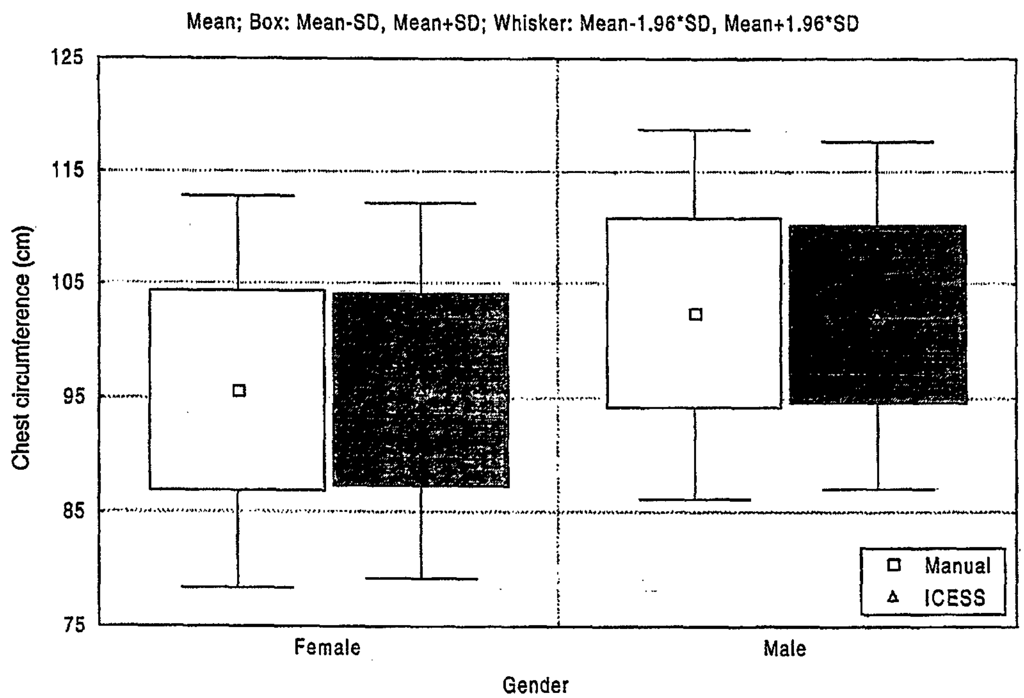


Figure 7 Comparison of manual and ICESS chest circumference measurements.

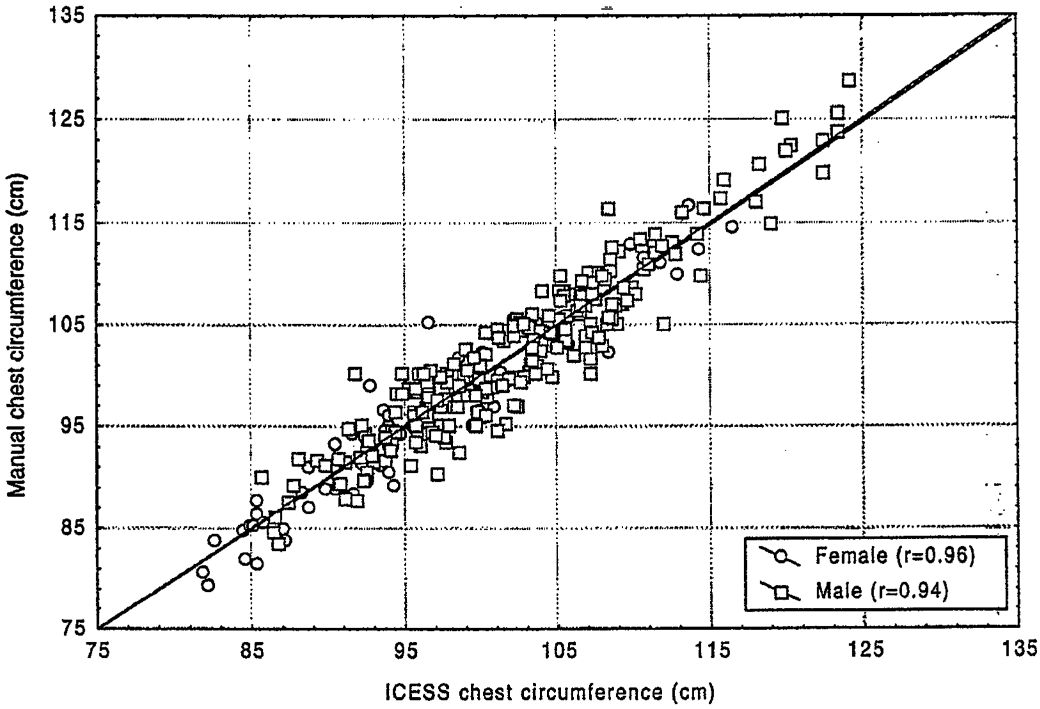

Figure 8 Scatter plot of manual and ICESS chest circumference measurements.

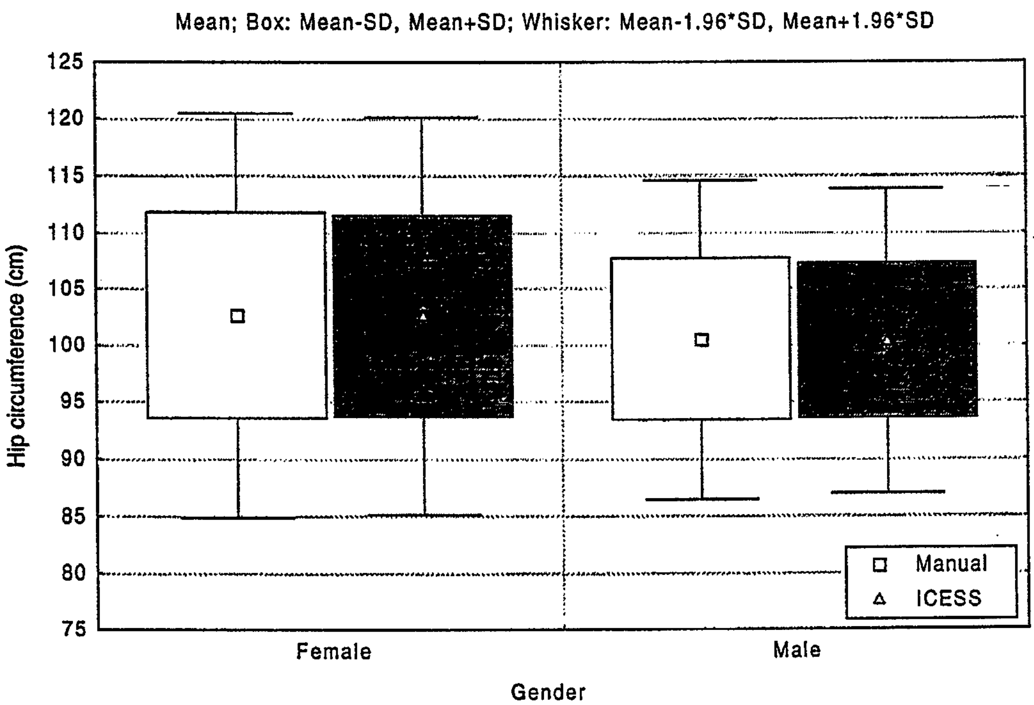

Figure 9 Comparison of manual and ICESS hip circumference measurements.

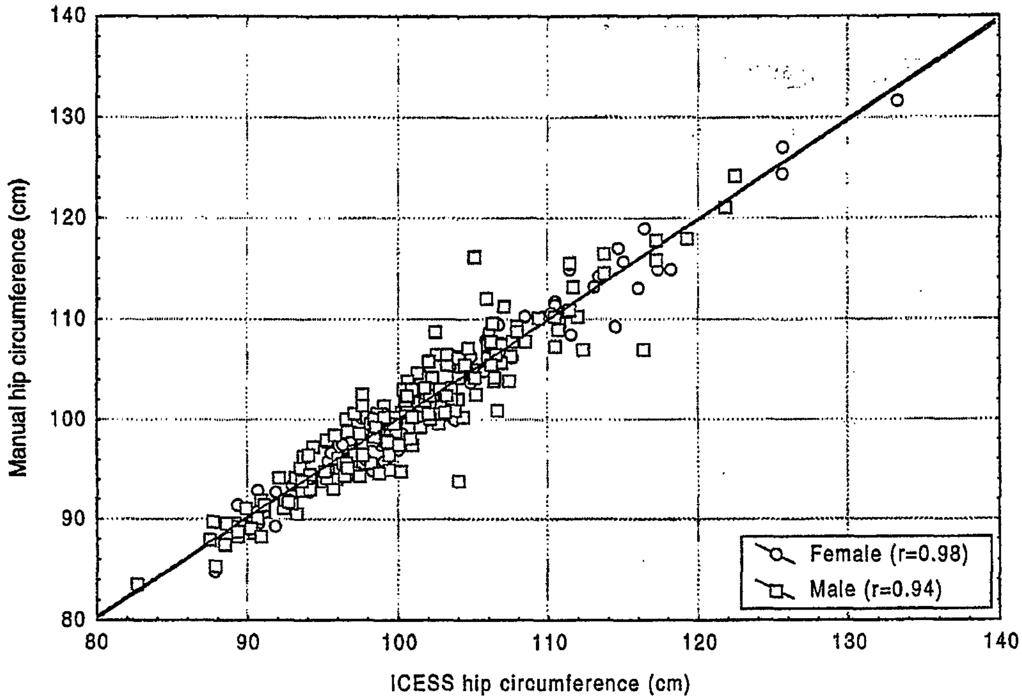


Figure 10 Scatter plot of manual and ICESS hip circumference measurements.

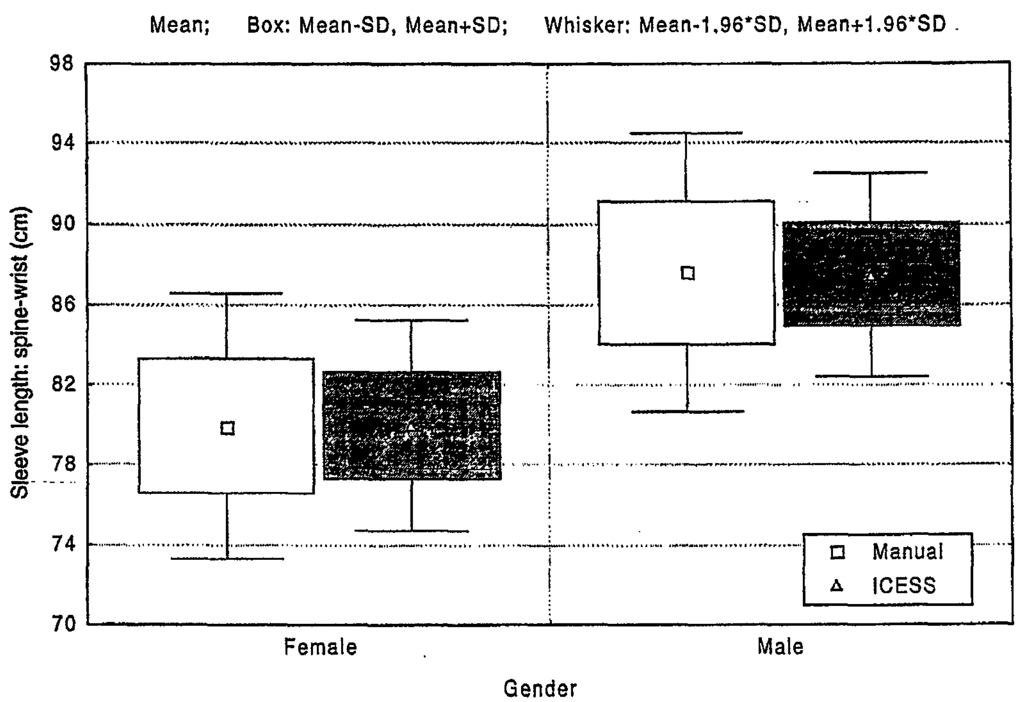


Figure 11 Comparison of manual and ICESS sleeve length measurements.

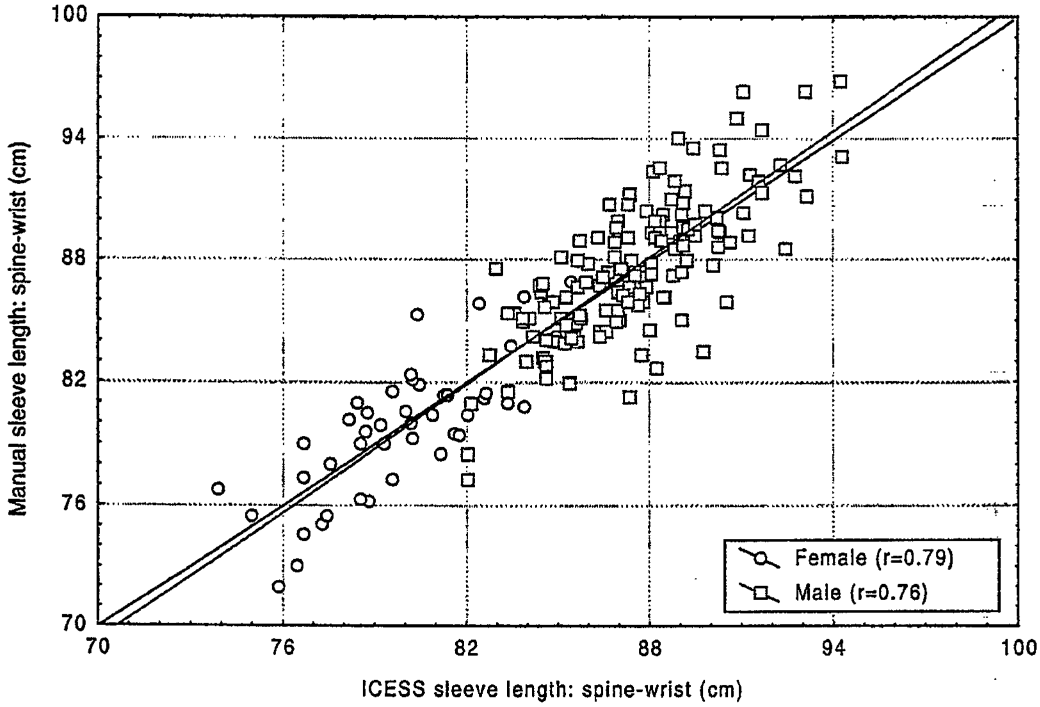


Figure 12 Scatter plot of manual and ICESS sleeve length measurements

## Precision

Mannequin tests

Thirty-five repeated measurements were made on a full size plastic mannequin. As before, the measurements analysed were those currently required for clothing sizing. All image capture and analysis sequences were performed in succession (every minute or so) such that camera calibration and lighting conditions were relatively constant. Table 1 summarises the results of this test. Interpretation of Table 1 shows that, for instance, an average stature of 182.20 cm was obtained; the range (difference between maximum and minimum values obtained in the test) was 0.27 cm; 68% of the measurements were within 0.068 cm of the mean (standard deviation (Std. Dev.)), while 95% were within 0.133 cm of the mean (1.96 Std. Dev. column).

Table 1 Mannequin repeatability results (cm).

| Variable | Mean | Range | Std.Dev. | 1.96 Std.Dev. |
|---|---|---|---|---|
| Stature | 182.20 | 0.27 | 0.07 | 0.13 |
| Neck circumference | 35.96 | 0.51 | 0.13 | 0.26 |
| Hip circumference | 94.65 | 1.24 | 0.32 | 0.63 |
| Waist circumference | 85.59 | 0.90 | 0.27 | 0.54 |
| Chest circumference | 95.98 | 1.28 | 0.31 | 0.61 |
| Sleeve length | 83.11 | 4.29 | 1.10 | 2.15 |

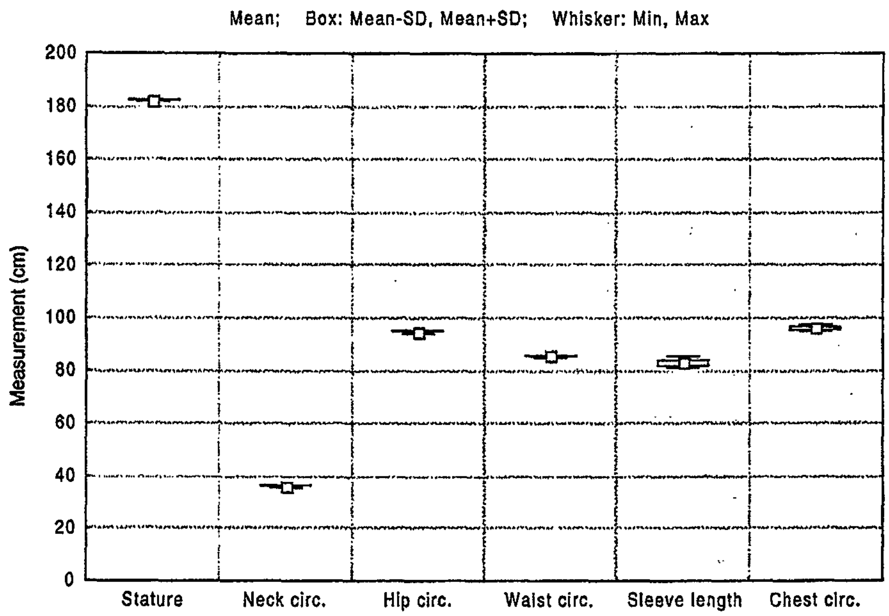


Figure 13 Box plot of repeated measurements of a mannequin with ICESS.

Human tests

Ten measurement cycles of a single individual were made within a 15-minute period. The results are shown in Table 2 and Figure 14.

Table 2 Human repeatability results (cm).

| Variable | Mean | Range | Std.Dev. | 1.96 Std. Dev. |
|---|---|---|---|---|
| Stature | 181.70 | 0.46 | 0.16 | 0.32 |
| Neck circumference | 36.87 | 0.58 | 0.19 | 0.38 |
| Hip circumference | 97.83 | 1.14 | 0.39 | 0.77 |
| Waist circumference | 87.33 | 1.51 | 0.49 | 0.95 |
| Chest circumference | 96.42 | 1.57 | 0.57 | 1.11 |
| Sleeve length | 88.70 | 3.56 | 1.02 | 2.01 |

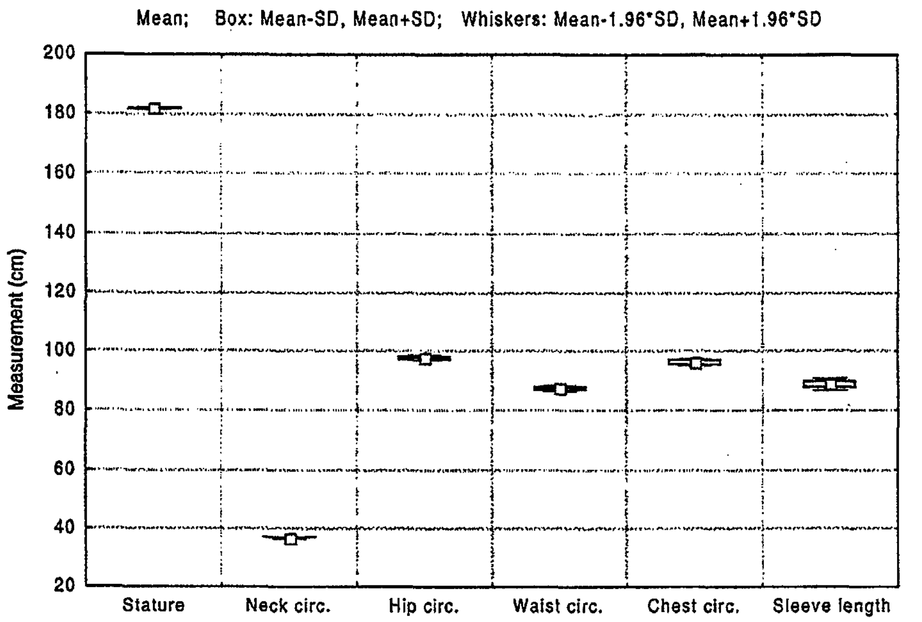


Figure 14 Box plot of repeated measurements of a male subject with ICESS.

## DISCUSSION

### Accuracy

As a group, the overall results did not indicate the presence of large systematic errors in ICESS compared to the manual measurements made during the LF97 survey. This is not surprising since indirect measurement models were fine-tuned using the LF97 survey data. However, there was evidence of differences with respect to the spread of results of sleeve length mostly, and to a lesser degree female neck circumference. In both cases, the spread of ICESS results was somewhat smaller than those taken manually (see Figures 5 and 11). The small difference in neck measurement spreads may have been due to differences in landmark identification and means of measurement between the two methods. In the manual method, accuracy can suffer from improper positioning of the measuring tape and skin compression. In image-based measurement, accuracy can suffer from unreliable landmarking and inadequate circumference modelling.

The difference in distributions is even greater for the sleeve length measurement (Figure 11). The accuracy of sleeve length suffers from variations in the posture of subjects, on one hand, and wrist and shoulder landmark detection inconsistencies. A study of the LF97 survey images confirmed the presence of inconsistent hand postures (some in pronation, some in supination), arms that were not in a vertical plane, and bent elbows. These can be remedied by providing subjects with better instructions on how to achieve the proper posture. In fact, since the survey, better control of posture has helped obtain results that were more consistent. Although the unreliability of landmark detection is partially remedied by adopting a proper posture, improvements to the algorithms will nevertheless be required in order to improve accuracy.

## Precision

The theoretical assessment of the measurement error made earlier suggested that an error of the order of ± 0.4 cm and ± 1.3 cm could be expected on direct and indirect measurements respectively. The results of repeatability tests performed on the plastic mannequin showed the actual errors to be smaller, indicating that the theoretical assumptions were perhaps a little too conservative. The direct measurements of stature were within 0.13 cm of the mean, 95% of the time. Where the mannequin's shape attributes were true to life (i.e. except for hinged joints, non-standard posture and unnatural shapes), reliable landmark positions were obtained. Hinges at the shoulder, elbow and wrist hindered the repeatability of sleeve length measurements. Fluctuations in this measurement in particular were unavoidable because the landmark detection software was developed to recognise real human shape. Other than for the neck, circumferences were found to be within 0.63 cm of the mean, 95% of the time (Table 1). Neck circumference exhibited significantly better repeatability due, in part, to special attention paid during the development and the fact that it is relatively easy to locate and measure.

Overall, it would appear that segmentation and landmark identification errors tend to fluctuate, on average, by one pixel on a given direct measurement, rather than the assumed two. The ratio of three between direct and indirect measurement error derived in the theoretical assessment was consistent with the circumference measurements observed in the data, i.e. $\pi$ x 1 pixel x 0.2 cm/pixel $\approxeq$ 0.63 cm.

For the most part, repeated measurements of a human subject showed the same basic trend as for the mannequin, i.e. direct measurements were more precise than circumferences, and neck circumference was more repeatable than other circumferences. In most cases, the human results exhibited more variability in measurement than the mannequin did. Figure 15 shows a comparison of the spread of measurements (1.96 x standard deviation) for both the mannequin and human subject. The largest difference between mannequin and human subject measurements are for waist and chest circumferences. This can be partly explained by torso movement during breathing (expansion and contraction of the rib cage and abdomen) and differences in posture from picture to picture (arm position, relaxed or tight posture).

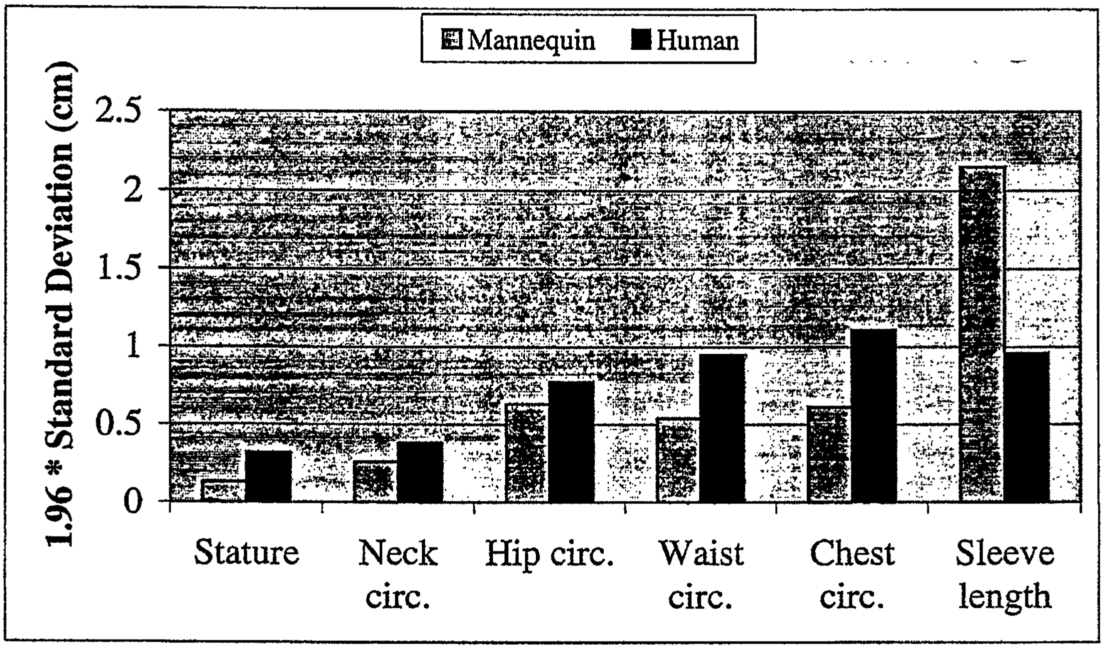


Figure 15 Range of measurements (1.96*SD) observed in repeated measures of a mannequin and human subject.

The results of the ICESS repeatability study on a human subject were compared with those of recent large-scale surveys where accuracy and precision were monitored throughout. The first survey was conducted on the Canadian Land Forces personnel in 1997 (Chamberland *et al.*, 1998). The second survey was conducted on US Army personnel in 1988 (Gordon *et al.*, 1989). In both cases, repeated measurements were implemented as part of the routine during the survey. The LF97 measurement error data pertains to a single observer repeating measurements on the same subject with the same landmarks within minutes (10 to 90 minutes) of the first measurement (see Forest *et al.*, 1999 for details). This can be viewed as the best case scenario in terms of repeatability, since it is assumed that the same observer will measure in the same way every time. The approach used in the US Army survey was similar in all respects except that that the re-measurement was done by a second observer. This case can be viewed as the best case scenario for repeatability by different observers, since both observers were highly trained on the dimensions specific to their measuring station.

The technical error of measurement (TEM), which is essentially a form of standard deviation, was used as the basis for comparison. Figure 16 shows the TEMs for ICESS measurements on a mannequin and human compared to single (Forest *et al.*, 1999) and dual observer results (taken from Gordon & Bradtmiller, 1992). The results indicate that the repeatability of ICESS measurements made on the mannequin and human are similar to the single observer results for stature and neck circumference. The single observer results had the lowest TEMs for all other measurements, followed by ICESS measurements on a mannequin and on a human. The TEM results of re-measurements made by two observers were worse than either of the ICESS TEMs. In its current configuration, ICESS repeatability of mannequin measurements is better than that obtained when two highly trained observers measure the same subject with the same landmarks, but slightly worse than when a single highly trained observer does the same. The differences observed between mannequin and human repeatability results show that the effect of posture and breathing during image capture is measurable by ICESS. Better precision could be obtained by controlling these factors, if required.

It should be noted that the survey results did not include landmarking error (the subjects had the same landmarks during re-measurement), whereas the ICESS results (the landmarks are located automatically after each image capture). Thus, if landmarking error were to be taken into consideration in the manual survey data, then ICESS would compare even more favourably.

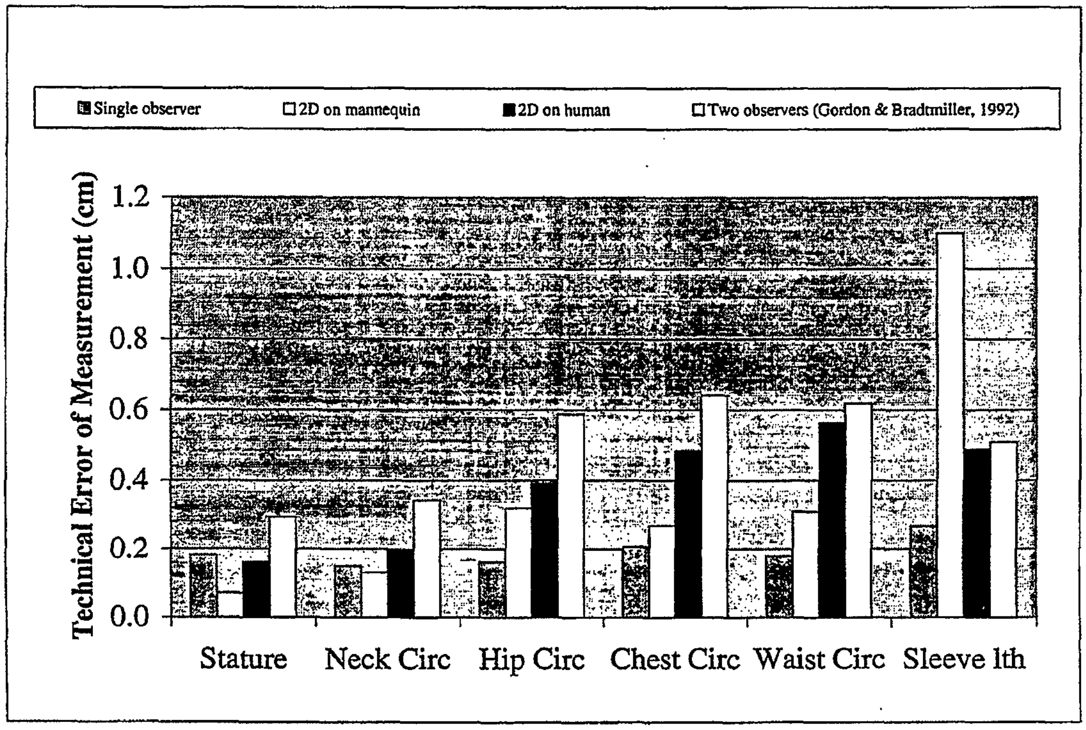


Figure 16 Comparison of TEM (technical error of measurement) obtained by ICESS on a human subject and expert manual measurements.

**Reliability**

Mueller & Martorell, 1988 state that two pieces of information are sufficient to characterise the reliability of an anthropometric variable: the TEM and the reliability coefficient. The reliability coefficient (R) is an interesting metric in that it compares the variability due to measurement error ($r^2$) against the biological variability of that dimension (sample variance $s^2$). It is computed using the following equation:

$$R = 1 - \left( \frac{r^2}{s^2} \right)$$

$$r = \sqrt{\frac{\sum_{i=1}^{n} \left( \sum_{j=1}^{k} x_j^2 - \frac{1}{k} \left( \sum_{j=1}^{k} x_j \right)^2 \right)_i}{n(k-1)}}$$

where r is the technical error of measurement, n is the number of subjects and k is the number of measurements per subject.

If the measurement error is small compared to the standard deviation of the sample then the reliability of that measurement will be high. Reliabilities above 90 to 95% have been recommended for the selection of variables in a survey (Gordon & Bradtmiller, 1992). The reliability coefficients obtained by ICESS were well above that for the dimensions shown in Table 3.

Table 3 Reliability of ICESS measurements for five anthropometric variables

| | Reliability |
|---|---|
| Stature | 99.9% |
| Neck circ. | 99.3% |
| Hip circ. | 99.7% |
| Chest circ. | 99.6% |
| Waist circ. | 99.7% |

**Clothing perspective**

The ultimate goal of ICESS is to determine the best fitting size of garment for a given individual. Anthropometry is one side of the equation, but clothing size and design is on the other. An idea of how much accuracy and precision is required for clothing size prediction can be obtained by considering the clothing itself. The following are a few of the factors that offer some clues as to how much accuracy is required. These are:

- Garment design or cut. If the clothing is more forgiving, i.e. is either loose fitting (such as combat clothing) or elastic (underwear), then a low degree of accuracy is all that is required. If the clothing is less forgiving, i.e. a close fitting dress uniform, then a higher degree of accuracy and precision is required, but only in key areas. Even in close fitting garments, there is a certain amount of ease included to allow for movement and comfort. Shirts are usually loose around the chest but snug at the neck, for instance.

- Manufacturing tolerance. It is difficult (and costly) to maintain tight manufacturing tolerances on manufactured items such as clothing. Table 4 shows some of the manufacturing tolerances currently in effect for CF trousers and shirts. While a high degree of accuracy and precision in anthropometric measurements is always desirable, it must be balanced against the ease provided in the garment design and the magnitude of manufacturing tolerances. The overall effectiveness of a clothing sizing system will only be as good as the weakest link.

Table 4 Manufacturing tolerances for CF dress trouser and shirt

| | | Tolerance (cm) |
|---|---|---|
| Trouser | waist | ± 1.3 |
| | inseam | ± 1.3 |
| Shirt | neck | ± 0.3 |
| | chest | ± 1.3 |
| | sleeve | ± 1.3 |

- Clothing size increments. The clothing size increments are an indicator of the criticality of some of the body measurements and of the importance given to fit. Clothing items that only require three sizes will either be very adjustable or very loose fitting. Consequently, accurate measurement of the body will not be necessary. Clothing items that require 40 sizes, such as in the case of the dress shirt, reflect the need to achieve good fit (and a lack of adjustability). Size increments for the dress uniform are shown in Table 5.

Table 5 Clothing size increments for CF dress uniform

| | | size increments (cm) |
|---|---|---|
| Trousers | stature | 7.6 |
| | waist | 5.1 |
| Shirt | neck | 1.3 |
| | sleeve length | 5.1 |
| Jacket | stature | 7.6 |
| | chest | 5.1 |

**Body variation**

Anthropometric accuracy and precision must also be balanced against body changes over minutes (breathing), hours (diurnal changes such as stature), days (weight changes), weeks (waist circumference changes), etc. Several body dimensions can change substantially over a short periods. Stature, for instance, has been known to change by 3 to 5 cm in a day depending on the amount of standing, walking and carrying done (NASA, 1978). In view of this type of fluctuation, it does not seem reasonable to measure within 0.1 cm a variable that can change by an order of magnitude during the course of the day. Stature to the nearest centimetre or so should be sufficient.

Davenport *et al.*, 1935 also reported changes in various body dimensions over time. In those experiments, repeated measurements of one subject were made at various times of day over a number of days by the same observer. The results (Table 6) show that measurements varied significantly. For waist circumference measurements, 95% of them were within ± 2.1 cm of the mean. Again, one could argue that measurement to within 0.1 cm is unnecessary for a dimension that can vary by an order of magnitude over a few days.

Table 6 Results of repeated measurements of a subject at various times of day over several days by one observer (Davenport *et al.*, 1935).

| | 1.96* s.d. (cm) |
|---|---|
| Waist circumference | 2.1 |
| Chest circumference | 1.5 |
| Neck circumference | 0.5 |

**Measurement accuracy requirements**

The first part of the discussion dealt with the capabilities of the image-based measurement system when compared with skilled human measurement. But the answer to the question "How much measurement accuracy is required?" can only be answered in the context of the application. For clothing sizing, a large part of the answer comes from the manufacturing tolerances. In a sense, the manufacturing tolerances represent the limits of a trade-off between fit of the clientele and cost of the garment. They could be interpreted as an amount of fluctuation in garment dimensions having minimal impact on fit for most of the customers of that nominal size. By extension, it could be said that given a garment size, the same amount of fluctuation in body measurement would also have minimal impact on the fit of a garment.

From a measurement standpoint, it is also important to balance the accuracy against short-term body variations. These variations, which occur naturally, must be accommodated by the clothing regardless of their magnitude in order for the clothing to be acceptable. Thus, using this argument, it would stand to reason that the magnitude of short-term body variations should temper measurement accuracy. A comparison of tables 4 and 6 shows a certain agreement between manufacturing tolerances and the short-term body variations that clothing must accommodate. Hence, it can be concluded that, in a balanced approach, measurement system accuracy should also be consistent with both. Therefore, from a practical perspective, neck circumference should be measured within ± 0.5 cm of the true value, whereas all other dimensions should be within ± 1.5 cm.

CONCLUSIONS

ICESS measurements were repeatable within 0.1 cm on stature and 0.6 cm on waist, hip, and chest circumferences 95% of the time, on a mannequin. Neck circumference was the most repeatable of circumference measurements, being within 0.3 cm of the mean 95% of the time.

From the analysis of short-term body changes, clothing design, fit, and manufacturing tolerances, it was clear that most dimensions used for clothing do not require a high degree of accuracy. It was determined that a body measurement system should be capable of measuring neck circumference within ± 0.5 cm in order to be effective, and all other dimensions within ± 1.5 cm. From the accuracy and precision analyses, it was concluded that the ICESS system was capable of these accuracies.

When properly designed and calibrated, image-based systems can provide unbiased anthropometric measurements that are quite comparable to traditional measurement methods (performed by skilled anthropometrists), both in terms of accuracy and repeatability. The quality of the results depends in large part on the dependability of the automatic landmarking algorithms and the correct modelling of the indirect measurements, but once this is achieved, this type of system can provide a reliable basis for the measurement of the CF population, regardless of where, when or by whom, it is operated.

## REFERENCES


Beckwith, T., Marangoni, R., & Lienhard V, J. (1993). *Mechanical measurements.* (5th ed.): Addison-Wesley.

Bennett, K., & Osborne, R. (1986). Interobserver measurement reliability in anthropometry. *Human Biology , 58*(5), pp. 751-759.

Chamberland, Ȧ., Carrier, R., Forest, F., & Hachez, G. (1998). *Anthropometric survey of the Land Forces (LF97)* (Contractor report 98-CR-15). Toronto, Ontario: Defence and Civil Institute of Environmental Medicine.

Davenport, C., Steggerda, M., & Drager, W. (1935). Critical examination of physical anthropometry on the living. *Proc. Amer. Acad. Arts Sci. Boston* (69), pp. 265-285.

Forest, F., Chamberland, A., Billette, J., & Meunier, P. (1999). *Anthropometric survey of the Land Forces: measurement error analysis.* (DCIEM No. CR 1999-041). Toronto: DCIEM.

Gordon, C. C., & Bradtmiller, B. (1992). Interobserver error in a large scale anthropometric survey. *American Journal of Human Biology* (4), pp. 253-263.

Gordon, C. C., Churchill, T., Clauser, C. E., Bradtmiller, B., McConville, J. T., Tebbetts , I., & Walker , R. A. (1989). *1988 anthropometric survey of US army personnel: methods and summary statistics.* (NATICK/TR-89/044). Natick, Ma: US Army Natick Research, Development, and Engineering Center.

Jamison, P., & Zegura, S. (1974). A univariate and multivariate examination of measurement error in anthropometry. *American Journal of Physical Anthropology* (40), pp. 197-204.

Kemper, H., & Pieters, J. (1974). Comparative study of anthropometric measurements of the same subjects in two different institutes. *American Journal of Physical Anthropology* (40), pp. 341-344.

Mueller, W., & Martorell, R. (1988). Reliability and accuracy of measurement. In A. R. a. R. M. TG Lohman (Ed.), *Anthropometric standardization reference manual* (pp. 83-86). Champaign, Illinois, USA: Human Kinetics.

NASA. (1978). *Anthropometry for designers*. (Vol. 1). Washington, DC: US Government printing office.

Schalkoff, R. (1989). *Digital Image Processing and Computer Vision*: Wiley.

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)

# DOCUMENT CONTROL DATA

(Security classification of title, body of abstract and indexing annotation must be entered when the overall document is classified)

| | |
|---|---|
| 1. ORIGINATOR (the name and address of the organization preparing the document. Organizations for whom the document was prepared, e.g., Establishment sponsoring a contractor's report, or tasking agency, are entered in section 12.)<br>DCIEM | 2. DOCUMENT SECURITY CLASSIFICATION (overall security classification of the document including special warning terms if applicable) |

3. DOCUMENT TITLE (the complete document title as indicated on the title page. Its classification should be indicated be the appropriate abbreviation (S,C,R or U) in parentheses after the title.)

Measurement accuracy and precision assessment of an image-based clothing and equipment sizing system

4. DESCRIPTIVE NOTES (the category of the document, e.g., technical report, technical note or memorandum. If appropriate, enter the type of report, e.g. interim, progress, summary, annual or final. Give the inclusive dates when a specific reporting period is covered.)
Report

5. AUTHOR(S) (Last name, first name, middle initial. If military, show rank, e.g. Burns, Maj. Frank E.)

P. Meunier and S. Yin

| | | |
|---|---|---|
| 6. DOCUMENT DATE (month and year of publication of document)<br>May 99 | 7.a. NO. OF PAGES (total containing information. Include Annexes, Appendices, etc.)<br>20 | 7.b. NO. OF REFS. (total cited in document)<br>12 |

| | |
|---|---|
| 8.a. PROJECT OR GRANT NO. (if appropriate, the applicable research and development project or grant number under which the document was written. Please specify whether project or grant)<br>6ke13 | 8.b. CONTRACT NO. (if appropriate, the applicable number under which the document was written) |
| 9.a. ORIGINATOR'S DOCUMENT NUMBER (the official document number by which the document is identified by the originating activity. This number must be unique to this document.)<br>DCIEM TR1999-070 | 9.b. OTHER DOCUMENT NO.(S) (any other numbers which may be assigned this document either by the originator or by the sponsor.) |

10. DOCUMENT AVAILABILITY (any limitation on further dissemination of the document, other than those imposed by security classification)

| | |
|---|---|
| x | Unlimited distribution |
| | Distribution limited to defence departments and defence contractors; further distribution only as approved |
| | Distribution limited to defence departments and Canadian defence contractors; further distribution only as approved |
| | Distribution limited to government departments and agencies; further distribution only as approved |
| | Distribution limited to defence departments; further distribution only as approved |
| | Other |

11. ANNOUNCEMENT AVAILABILITY (any limitation to the bibliographic announcement of this document. This will normally correspond to the Document Availability (10.) However, where further distribution (beyond the audience specified in 10) is possible, a wider announcement audience may be selected.)

12. SPONSORING ACTIVITY (the name of the department project office or laboratory sponsoring the research and development. Include the address.)
Simulation and Modelling in Acquisition, Rehearsal and Training
Defence and Civil Institute of Environmental Medicine
PO Box 2000, 1133 Sheppard Avenue West
North York, Ontario, M3M 3B9

DSIS DCD03
HFD 09/94

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)

UNCLASSIFIED

SECURITY CLASSIFICATION OF FORM
(Highest classification of Title, Abstract, Keywords)

13. ABSTRACT ( a brief and factual summary of the document. It may also appear elsewhere in the body of the document itself. It is highly desirable that the abstract of classified documents be unclassified. Each paragraph of the abstract shall begin with an indication of the security classification of the information in the paragraph (unless the document itself is unclassified) represented as (S), (C), (R), or (U). It is not necessary to include here abstracts in both official languages unless the text is bilingual).

In spite of highly standardised protocols designed to maximise the degree of repeatability and accuracy, anthropometric data are not always as reliable as they appear. Many factors come into play during the measurement of human subjects, resulting in numerous possible sources of error. Researchers have found the magnitude of these errors to be such that, even if measured by highly trained observers, comparison of two populations may be meaningless.

Computerised image-based systems can overcome some of the problems of traditional anthropometry, such as error due to the pressure exerted on soft tissue by measurement instruments, or transcription errors. However, all sources errors are not eliminated by computerised measurement. In image-based systems, the sources of error take the form of perspective distortion, camera resolution, and inadequacy of the mathematical models used to estimate circumference measurements.

The accuracy of measurements made by an image-based clothing and equipment sizing system was estimated using a database of 349 subjects (male and female) who were also measured traditionally. The precision, or repeatability, of this system was estimated through repeated measurements of both a plastic mannequin and a human. Although the image-based system did not exhibit systematic bias in the results, the standard deviations were somewhat smaller for some dimensions than those obtained by manual measurement. The repeatability results were comparable to those obtained by highly trained anthropometrists, as reported in recent large-scale surveys. The reliability of the measurements needed for clothing, i.e. the proportion of error of measurement to biological variability, was greater than 99% in all cases.

The degree of accuracy and precision of the measurements required for the selection of clothing and equipment size was put into perspective with the realities of short-term fluctuations in body size, clothing design, and manufacturing tolerances. When a balanced approach is used, neck circumference is found to be, by far, the anthropometric dimension requiring the greatest amount of accuracy. Because of the ease with which it can be identified and measured by image processing, it is also the system's most accurately measured circumference.

When properly designed and calibrated, image-based systems can provide unbiased anthropometric measurements that are quite comparable to traditional measurement methods (performed by skilled measurers), both in terms of accuracy and repeatability. The quality of the results depends, in large part, on the dependability of the automatic landmarking algorithms and the correct modelling, but once this is achieved, this type of system can provide a reliable basis for the measurement of a population, regardless of where, when or by whom, it is operated.

14. KEYWORDS, DESCRIPTORS or IDENTIFIERS (technically meaningful terms or short phrases that characterize a document and could be helpful in cataloguing the document. They should be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location may also be included. If possible, keywords should be selected from a published thesaurus, e.g. Thesaurus of Engineering and Scientific Terms (TEST) and that thesaurus identified. If it is not possible to select indexing terms which are Unclassified, the classification of each should be indicated as with the title.)

ICESS, photogrammetry, anthropometry, measurement error

The Defence Research
and Development Branch
provides Science and
Technology leadership
in the advancement and
maintenance of Canada's
defence capabilities.

Leader en sciences et
technologie de la défense,
la Direction de la recherche
et du développement pour
la défense contribue
à maintenir et à
accroître les compétences
du Canada dans
ce domaine.

#511502

DEFENCE R&D DÉFENSE

www.crad.dnd.ca